Interoperability Issues

Ross Overbeek,
Fellowship for Interpretation of
Genomes
(FIG)

www.nmpdr.org

Basic Data Types

- Contigs
- Gene Calls
- Function Assignments
- Annotations (of genes)
- Protein families
- Metabolic reconstructions/subsystems
  - stoichiometric matrices
- Structure assessments
- Expression Data
- Characterization of variation among a set of closely-related genomes
- Correlations of variation with phenotype/measurement

www.nmpdr.org

NMPDR
National Microbial
Pathogen Data
Resource Center

Central Issues?

1. Can we agree on what data should be offered?

2. Can we offer it in a consistent manner?

3. Can we form effective relationships with external groups to synchronize maintenance of data?

   - NCBI

   - UniProt

   - KEGG

Issues

- Contigs
  - Representation of quality
- Gene Calls
  - What feature types should be supported?
  - Consistency among a set of genomes
- Function Assignments
  - Can we use/develop a common Gene Ontology?
- Annotations
  - There are structured and unstructured annotations.
  - Can we rapidly establish a common set of structured annotations?

NMPDR
National Microbial
Pathogen Data
Resource Center

Issues (continued)

- Protein families

  – There are alternatives

  – We should link to all

  – Does it make sense to coordinate development of the tools to maintain links?

NMPDR
National Microbial
Pathogen Data
Resource Center

Issues (continued)

- Metabolic reconstructions/subsystems
  - Beyond an ontology, does it make sense to coordinate these efforts?

- Stoichiometric matrices
  - We consider these desirable, but probably a topic to be deferred

NMPDR
National Microbial
Pathogen Data
Resource Center

Issues (continued)

- Structure Assessments
  - Is it possible to establish a coordinated development of links?

- Expression Data
  - Does it make sense to define a core set of services and attempt a consistent implementation?

NMPDR
National Microbial
Pathogen Data
Resource Center

Issues (continued)

- Characterization of Variation
  - This is a central capability that will grow in significance
  - We should seek a consistent representation
    - Major variation (rearrangements, insertions, deletions)
    - Minor variation (SNPs)
  - Tools to correlate variation and phenotype/measurement are in an intial state

NMPDR
National Microbial
Pathogen Data
Resource Center

Specific Recommendations

- Use DAS to exchange annotations and to present a common access framework

- Develop an ontology and use it

- Coordinate development of a few key tools that everyone will use

  - characterization of variation among a closely related set of genomes

  - mining literature

- Support web services. Define a set of core capabilities that everyone supports